# CHEMOINFORMATICS AND ITS APPLICATIONS

***Wenli XU***
***Min LING***
***Jing HU***
***Ying HUANG***
***Jia LI, PhD***
***Jianhua YAO , PhD, Professor***
Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, China

**Abstract:**

With development of computer science and network, computer and software are getting more and more applications in scientific research, production and service areas. *In silico* is an expression used to mean "performed on computer or via computer simulation".

Chemoinformatics is one of *in silico* technologies employed in researches relating to Chemistry. It is an application of informatics methods to solve chemical problems concerned with molecular design, synthesis design and structural identification. It is an interdisciplinary combining of Computer science, Mathematics and Chemistry. In principle, three methods are included in Chemoinformatics: data-based, logic-based and principle-based.

These years, Chemoinformatis has been applied not only in chemical research but also in domains relating to Chemistry, i.e., Drug Discovery, Pesticide Design, Environment Protection, Material Design, Traditional Chinese Medicine (TCM), Food Safety and etc., which relate to Chemistry. Chemoinformatics will be applied in more and more domains relating to Chemistry because "Chemistry is (almost) everywhere and in everything".

Herein, two works: 1. modern mode of chemical research or other works relating to chemical research is proposed, which is combined by thinking, experiment and *in silico*; 2. *in silico* platform for pesticide design will be presented.

**Key Words:** Chemoinformatics, *In silico*, Chemistry

## Introduction

Chemistry is a branch of physical science and focused on studying the composition, properties and behavior of matter. In general, it includes 23 sub disciplines: 1) Acid-base reaction theories; 2) Alchemy; 3) Analytical chemistry; 4) Astrochemistry; 5) Biochemistry; 6) Crystallography; 7) Environmental chemistry; 8)Food chemistry; 9) Geochemistry; 10) Green chemistry; 11) Inorganic chemistry; 12) Materials science; 13) Molecular physics; 14) Nuclear chemistry; 15) Organic chemistry; 16) Photochemistry; 17) Physical chemistry; 18) Radiochemistry; 19) Solid-state chemistry; 20) Stereochemistry; 21) Supramolecular chemistry; 22) Surface science and 23) Theoretical chemistry. [1] And it is (almost) everywhere and in everything [2]. In Chemistry, the three subjects: molecular modeling, synthesis design and structural identification, are considered as three chemical problems.

Chemoinformatics (also termed as cheminformatics and chemical informatics) is an application of informatics methods to solve the chemical problems concerned with molecular design, synthesis design and structural identification [3]. It is an interdisciplinary [4] combining of Computer science, Mathematics and Chemistry. In principle, three methods are included in Chemoinformatics: data-based, logic-based and principle-based. They are used at different stage and different levels in chemical study.

Data-based means that people retrieve data from database. It is usually employed at the beginning of study.

Logic-based means that people retrieve knowledge from normalized data and do predictions based on knowledge. In general, it is employed to decide if compounds should be synthesized or extaction.

Principle-based means that people do the study by quantum chemical calculation. The calculation results are often used in mechanism study.

Herein, two works: 1. modern mode of chemical research or other works relating to chemical research is proposed, which is combined by thinking, experiment and *in silico*; 2. *in silico* platform for pesticide design will be presented.

**Modern Mode of Chemical Research**

According to the history of Chemistry, ancient Egyptians pioneered the art of synthetic "wet" chemistry up to 4,000 years ago; ancient people did works relating to Chemistry by 1000 BC. [1] In chemical research, computer started to be used in1940s. At present, most chemists prefer traditional mode: thinking based on experiences and experiments (synthesis, extraction or separation).

The modern mode of chemical research showed in Figure 1, is development of traditional mode and combined by thinking, experiment and *in silico*. *In silico* is an expression used to mean "performed on computer or via computer simulation" [5]. Chemoinformatics is one of *in silico* technologies employed in researches relating to Chemistry.

In chemical researches, molecular design, synthesis design and structural identification, are often considered.

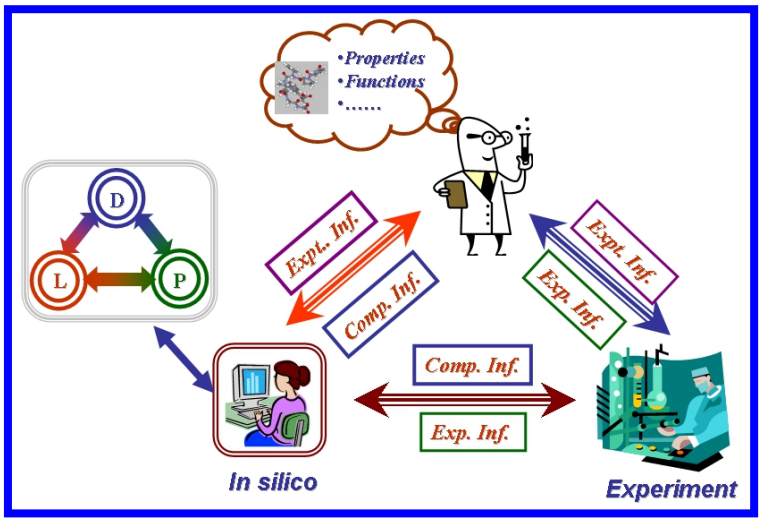


**Figure 1.** Modern Mode of Chemical Research

Molecular design is the application of all techniques leading to the discovery of new chemical entities with specific properties required for the intended application [6]. In traditional mode, people do molecular design and synthesis design mainly based on experiences and experiments. However, in modern mode, people do molecular design and synthesis design not only based on experiences and experiments, but also on database, knowledge base and quantum chemical calculation. The process of molecular design (showed in Figure 2) can be expressed as: 1. Inspiration (proposing a new chemical entity with specific properties based on experiences); 2. Virtual design (Search information in databases; Prediction of properties based on chemical structures) 3. Synthesis or extraction.

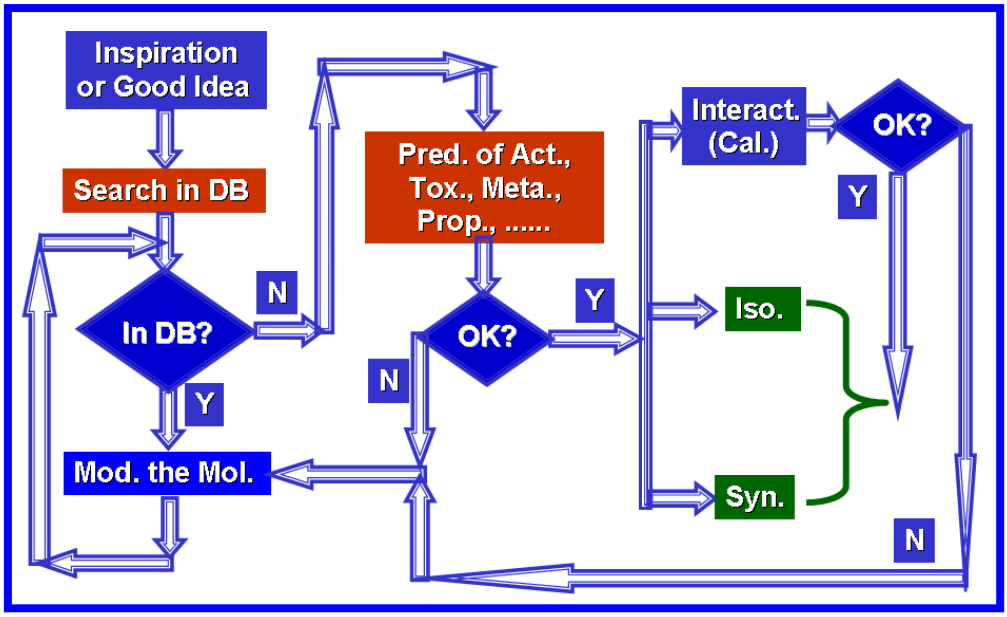


**Figure 2.** Flowchart of Molecular Design (*in silico*)

The work flow of synthesis is: 1. proposing a product; 2. deducing its candidate reagents based on retrosynthetic analysis [7]; 3. repeating 2. The flow will be stopped when the reagents are available.

**Chemoinformatics Platform for Pesticide Design**

The pesticide was defined by Food and Agriculture Organization (FAO) in 2002 and its definition is as followed: Pesticide means any substance or mixture of substances intended for preventing, destroying or controlling any pest, including vectors of human or animal disease, unwanted species of plants or animals causing harm during or otherwise interfering with the production, processing, storage, transport or marketing of food, agricultural commodities, wood and wood products or animal feedstuffs, or substances which may be administered to animals for the control of insects, arachnids or other pests in or on their bodies. The term includes substances intended for use as a plant growth regulator, defoliant, desiccant or agent for thinning fruit or preventing the premature fall of fruit, and substances applied to crops either before or after harvest to protect the commodity from deterioration during storage and transport [8]. It means that pesticide is related to Chemistry, environment, human health and etc. In pesticide design, molecular design is one part of most important works. Bioactivity, toxicity and reactivity of compounds which will become pesticide should be paid attention.

*In silico* platform (showed in Figure 3) for pesticide design has been applied in pesticide design.

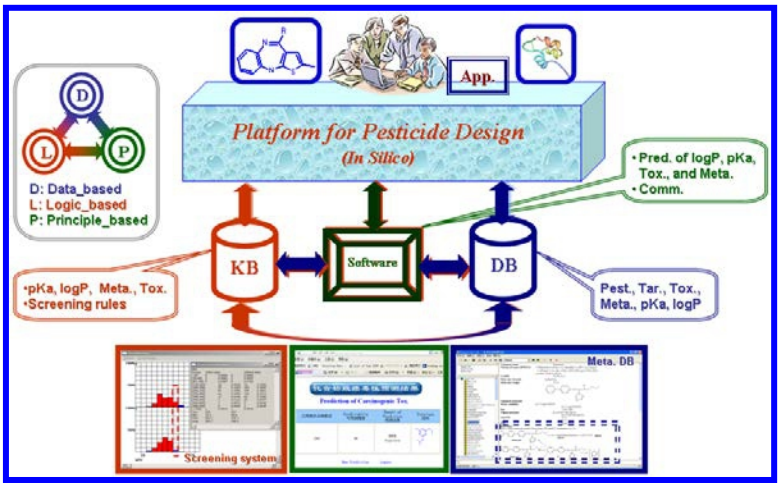


**Figure 3.** *In silico* Platform for Pesticide Design

This platform is supported by software, database and knowledge base as followed: 1) Prediction System of Carcinogenic Toxicity (CISOC-PSCT) [9, 10] is used to predict carcinogenic toxicity of compounds based chemical structures; 2) Prediction System of Mutagenic Toxicity (CISOC-PSMT) [11, 12] is used to predict mutagenic toxicity of compounds based chemical structures;; 3) Prediction System of Acute Toxicity"(CISOC-PSAT)[13] is used to predict acute toxicity of compounds based chemical structures; 4) Prediction System of logP (CISOC-logP) [14, 15] is used to predict the ratio of concentrations of a compound in a mixture of two immiscible phases at equilibrium according to the chemical structure. The ratio is the measure of hydrophilic ("water-loving") or hydrophobic ("water-fearing") for a chemical substance. 5) Prediction System of pKa"(CISOC-pKa) [16] is used to predict the acidity constant, or acid-ionization constant of a compound. The constant is a quantitative measure of the strength of an acid in solution; 6) Metabolism Knowledge Base Management System (CISOC-MetaKBMS) [17] is used to retrieve metabolism knowledge from normalized metabolism data. And the knowledge in the base can be used to predict metabolites of a chemical substance; 7) Receptor Information Search System (CISOC-WEBPESTRMS) [18] is used to search receptor or ligand information; 8) Web-based Agriculture Chemicals Database (in house) [19] is used to search information about commercial pesticides.

Before synthesis or extraction, scientists propose a compound, then they browse information in databases, do bioactivity, toxicity and reactivity predictions by the prediction systems. Quantum chemical calculation results are used to study chemical behaviors. They will have a list of candidates based on predictions, calculations and experiences.

**Conclusion**

The two examples mentioned in the front show: 1) the role of *in silico* in chemical research and pesticide design: increasing efficiency of work, decreasing pollution and costs; 2) people should pay attention on study and application of interdisciplinary.

**Acknowledgement**

This work was financially supported by MOST (973 Program, No: 2010CB126103; Key Program, No: 2011BAE06B05) and National Natural Science Foundation of China ( No: 21072216).

**References:**

[1] http://en.wikipedia.org/wiki/Chemistry.
[2] Arnon Shani, *Chem. Edu. Inter.*, Vol. 4, No. 1, 2003.
[3] J., Gasteiger, T., Engel (Eds): Chemoinformatics: A Textbook. John Wiley & Sons, 2004.
[4] http://en.wikipedia.org/wiki/Interdisciplinary.
[5] http://en.wikipedia.org/wiki/In_silico.
[6] Compendium of Chemical Terminology Gold Book, Version 2.3.2, 2012-08-19, International Union of Pure and Applied Chemistry, pp950.
[7] E. J., Corey; X-M., Cheng, The Logic of Chemical Synthesis. New York: Wiley, 1995.
[8] International Code of Conduct on the Distribution and Use of Pesticides, Food and Agriculture Organization of the United Nations, Rome, 2002.
[9] Quan. Liao, Jianhua Yao,et al, SAR and QSAR in Environmental Research, 2004, 15(3), 217-235.
[10] CISOC-PSCT, Registration Number (China): 022239.
[11] Quan Liao, Jianhua Yao, et al, Molecular Diversity, 2007, 11, 59-72.
[12] CISOC-PSMT, Registration Number (China): 042160.
[13] CISOC-PSAT V1.0, Registration Number (China): 0232654.
[14] Quan Liao, Jianhua Yao, et al, Molecular Diversity, 2006, 10, 301-309.
[15] CISOC-LogP, Registration Number (China): 046394.
[16] pKa prediction system (CISOC-pKa)V1.0, Registration Number (China): 0154688.
[17] CISOC-MetaKBMS V1.0, Registration Number (China): 0154319.
[18] CISOC-WEBPESTRMS V1.0, Registration Number (China): 0411338.
[19] http://202.127.145.116/pesticide/index.php.